第27卷 第4期　　工 程 数 学 学 报　　Vol. 27 No. 4
2010年08月　　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　　Aug. 2010

文章编号:1005-3085(2010)04-0637-06

一类新的记忆梯度法及其收敛性*

汤京永[1,2]，董　丽[1]

(1- 信阳师范学院数学与信息科学学院，信阳　464000；2- 上海交通大学数学系，上海　200240)

摘　要: 本文着重研究求解无约束优化问题的记忆梯度法，利用当前和前面一步迭代点的信息产生下降方向，采用Armijo线性搜索确定步长，得到了一类新的无约束优化算法。新算法在较弱的条件下具有全局收敛性和线性收敛速率，并且不用计算和存储矩阵，适于求解大规模优化问题。数值试验表明算法是有效的。

关键词: 无约束优化；记忆梯度法；全局收敛性；线性收敛速率

分类号: **AMS**(2000) 90C30; 65K05; 49M37　　**中图分类号**: O221.2　　**文献标识码**: A

1　引言

考虑无约束优化问题

$$\min \quad f(x), \quad x \in R^n, \tag{1}$$

其中R^n是n维欧氏空间，$f(x): R^n \longrightarrow R$为连续可微函数，$g(x)$为其梯度。求解问题(1)的算法主要是迭代法，基本结构为

$$x_{k+1} = x_k + \alpha_k d_k,$$

其中d_k为搜索方向，α_k为步长参数。对α_k和d_k的不同选择就构成了不同的迭代法[1]。在本文中，若x_k为当前迭代点，则简记$g(x_k)$为g_k，$f(x_k)$为f_k，$f(x^*)$为f^*。

共轭梯度法在每步迭代中不需计算和存储矩阵，是求解大型无约束优化问题的有效算法之一。记忆梯度法类似于共轭梯度法，也是求解大规模优化问题的有效方法，并且与共轭梯度法相比，记忆梯度法增加了参数选择的自由度，更有利于构造稳定的快速收敛算法[2-8]。

本文提出一类新的求解无约束优化问题的记忆梯度法，算法利用当前和前面一步迭代点的信息以及Armijo步长规则产生新的迭代点，结构简单，不用计算和存储矩阵，适于求解大规模优化问题。与文献[5-8]中的算法相比，新算法有下面三个优点：一是假设条件较弱，扩大了算法求解问题的范围；二是在较弱的条件下具有全局收敛性和线性收敛速率；三是具有较好的数值试验结果。

2　记忆梯度法及其性质

本文作如下假设：

(H_1)　目标函数$f(x)$在水平集$L_0 = \{x \in R^n \,|\, f(x) \leq f(x_1)\}$上有下界。

(H_2)　梯度函数$g(x)$在包含L_0的开凸集B上一致连续。

收稿日期: 2008-08-15.　**作者简介**: 汤京永(1979年6月生)，男，博士，讲师. 研究方向：优化理论与算法.

*__基金项目__: 国家自然科学基金(10571109)；信阳师范学院青年基金(200946; 200947).

算法1 取$\eta\in(\frac{1}{2},1)$，$\rho\in(0,1)$，$\beta\in(0,1)$，$x_1\in R^n$。令$k:=1$。

步骤1 若$\|g_k\|=0$，则停止迭代；否则，转步骤2。

步骤2 计算d_k，使其满足

$$d_k=\begin{cases}-g_k, & k=1,\\ -g_k+\beta_k\delta_{k-1}, & k\geq 2,\end{cases}\tag{2}$$

其中$\delta_{k-1}=d_{k-1}-g_{k-1}$，而

$$\beta_k=\begin{cases}0, & \text{若}\ d_{k-1}=g_{k-1},\\ \frac{\eta\|g_k\|}{\|\delta_{k-1}\|}, & \text{若}\ d_{k-1}\neq g_{k-1}.\end{cases}\tag{3}$$

步骤3 $x_{k+1}=x_k+\alpha_k d_k$，其中α_k由Armijo搜索准则确定，即要求α_k为$\{1,\beta,\beta^2,\cdots\}$中满足下式的最大者

$$f(x_k)-f(x_k+\alpha d_k)\geq -\rho\alpha g_k^T d_k.\tag{4}$$

步骤4 令$k:=k+1$，转步骤1。

引理1 对任意的$k\geq 1$，有$-g_k^T d_k\geq(1-\eta)\|g_k\|^2$。

注 由引理1及(4)式可知$\{f_k\}$单调不增且$x_k\in L_0$。

引理2 对任意的$k\geq 1$，有$\|d_k\|\leq(1+\eta)\|g_k\|$。

3 算法的全局收敛性

定理1 假设(H_1)和(H_2)成立，则算法1或有限步终止于问题(1)的稳定点；或产生无穷点列$\{x_k\}$，其任意聚点都是问题(1)的稳定点。

证明 若$g(x_k)=0$，则x_k为稳定点。假设算法1产生无穷点列$\{x_k\}$，x^*为其任意聚点，则存在子列$\{x_k:k\in K\}$，$K\subset\{1,2,\cdots\}$，使$\lim\limits_{k\in K,k\to\infty}x_k=x^*$。下面分两种情况讨论。

情况1 $\inf\limits_{k\in K}\alpha_k>0$。由(4)式和引理1可知

$$f_k-f_{k+1}\geq -\rho\alpha_k g_k^T d_k\geq\rho(1-\eta)\alpha_k\|g_k\|^2.\tag{5}$$

因为$\{f_k\}$单调不增有下界，故$\{f_k\}$有极限。由(5)式知$\lim\limits_{k\to\infty}\alpha_k\|g_k\|^2=0$，故

$$\lim_{k\in K,k\to\infty}\alpha_k\|g_k\|^2=0,$$

进而由$\inf\limits_{k\in K}\alpha_k>0$可知$\lim\limits_{k\in K,k\to\infty}\|g_k\|^2=0$，即

$$\lim_{k\in K,k\to\infty}\|g_k\|=0.$$

因为$\lim\limits_{k\in K,k\to\infty}x_k=x^*$，且$g(x)$连续，所以$g(x^*)=0$，故知$x^*$是问题(1)的稳定点。

情况2 $\inf\limits_{k\in K}\alpha_k=0$。由$\inf\limits_{k\in K}\alpha_k=0$知存在子列$N\subset K$，使

$$\lim_{k\in N,k\to\infty}\alpha_k=0,$$

故知存在$K' > 0$，当$k \in N$且$k > K'$时有$\alpha_k < 1$，进而可知$\alpha_k\beta^{-1} \in \{1,\beta,\beta^2,\cdots\}$。令$\alpha = \alpha_k\beta^{-1}$，则由Armijo搜索准则可得

$$f_k - f\big(x_k + \alpha_k\beta^{-1}d_k\big) < -\rho\alpha_k\beta^{-1}g_k^Td_k, \quad k \in N, \quad k > K'. \tag{6}$$

对不等式(6)的左边利用中值定理，则存在$\theta_k \in (0,1)$，使

$$-\alpha_k\beta^{-1}g\big(x_k + \theta_k\alpha_k\beta^{-1}d_k\big)^Td_k < -\rho\alpha_k\beta^{-1}g_k^Td_k, \quad k \in N, \quad k > K',$$

从而知

$$g\big(x_k + \theta_k\alpha_k\beta^{-1}d_k\big)^Td_k > \rho g_k^Td_k, \quad k \in N, \quad k > K'. \tag{7}$$

由Cauchy-Schwartz不等式，引理1及(7)式可知

$$\begin{aligned}\|g(x_k + \theta_k\alpha_k\beta^{-1}d_k) - g_k\| &\geq \frac{(g(x_k + \theta_k\alpha_k\beta^{-1}d_k) - g_k)^Td_k}{\|d_k\|}\\ &\geq \frac{(1-\rho)(1-\eta)\|g_k\|^2}{\|d_k\|}, \quad k \in N, \quad k > K',\end{aligned}$$

从而由引理2知

$$\|g(x_k + \theta_k\alpha_k\beta^{-1}d_k) - g_k\| \geq \frac{(1-\rho)(1-\eta)\|g_k\|}{1+\eta}, \quad k \in N, \quad k > K'. \tag{8}$$

因为$\{x_k : k \in K\}$有界，$g(x)$连续，故$\{\|g_k\| : k \in K\}$有界，从而由引理2可知$\{\|d_k\| : k \in K\}$有界。又因为$\lim\limits_{k\in N,k\to\infty} \alpha_k = 0$，故

$$\lim_{k\in N,k\to\infty} \alpha_k\|d_k\| = 0,$$

从而由(H_2)和(8)式知$\lim\limits_{k\in N,k\to\infty} \|g_k\| = 0$。因为

$$\lim_{k\in N,k\to\infty} x_k = x^*$$

且$g(x)$连续，所以有$g(x^*) = 0$，故知x^*是问题(1)的稳定点。

4 算法的线性收敛速率

假设：

(H_3) $f(x)$是二次连续可微的一致凸函数。

引理3[7] 若(H_3)成立，则$f(x)$在R^n上有唯一的极小点x^*，并且存在$M \geq m > 0$，使

$$\frac{m}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{M}{2}\|x - x^*\|^2, \tag{9}$$

$$m\|x - x^*\| \leq \|g(x)\| \leq M\|x - x^*\|. \tag{10}$$

引理4[7] 若(H_3)成立，则$g(x)$在水平集L_0上Lipschitz连续，即存在常数$L > 0$，使

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall\, x, y \in L_0.$$

定理2 若 (H_3) 成立，$\{x_k\}$ 是由算法1产生的无穷点列，则 $\{x_k\}$ 至少线性收敛于 x^*。

证明 若 $\alpha_k = 1$，则由 (5) 式可知

$$f_k - f_{k+1} \geq \rho(1-\eta)\|g_k\|^2. \tag{11}$$

若 $\alpha_k < 1$，则由定理1的证明过程可知，存在 $\theta_k \in (0,1)$，使

$$\left\|g(x_k + \theta_k\alpha_k\beta^{-1}d_k) - g_k\right\| \geq \frac{(1-\rho)(1-\eta)\|g_k\|}{1+\eta}. \tag{12}$$

由引理4及 (12) 可得

$$\alpha_k\beta^{-1}L\|d_k\| \geq \left\|g(x_k + \theta_k\alpha_k\beta^{-1}d_k) - g_k\right\| \geq \frac{(1-\rho)(1-\eta)\|g_k\|}{1+\eta},$$

从而由引理2知

$$\alpha_k \geq \frac{(1-\rho)(1-\eta)}{L(1+\eta)\beta^{-1}}\frac{\|g_k\|}{\|d_k\|} \geq \frac{(1-\rho)(1-\eta)}{L(1+\eta)^2\beta^{-1}}. \tag{13}$$

由 (5) 式和 (13) 式可得

$$f_k - f_{k+1} \geq \frac{\rho(1-\rho)(1-\eta)^2}{L(1+\eta)^2\beta^{-1}}\|g_k\|^2. \tag{14}$$

令

$$\omega = \min\left\{\rho(1-\eta), \frac{\rho(1-\rho)(1-\eta)^2}{L(1+\eta)^2\beta^{-1}}\right\},$$

则由 $0 < \rho < 1$，$\frac{1}{2} < \eta < 1$ 可知 $0 < \omega < \frac{1}{2}$。由 (11) 及 (14) 知 $f_k - f_{k+1} \geq \omega\|g_k\|^2$。

余下的证明类似于文献 [7] 中的定理3，故略。

5 数值试验

为进一步检验算法1的实算效果，我们选取几个算例对本文算法进行数值试验，并与Armijo搜索下的共轭梯度法和最速下降法进行比较。用IT表示算法的迭代次数，NA表示本文提出的新算法，用FR, PRP和HS分别表示FR, PRP和HS共轭梯度法[6]，SM表示最速下降法。参数取值为 $\eta = 0.88$，$\rho = 0.75$，$\beta = 0.5$。表中的数字为迭代中的目标函数值，舍入成小数点后有三位有效数字，一维搜索全部采用Armijo搜索。记 2.381×10^5 为2.381(5)，3.1×10^{-3} 为3.100(-3)，其余同。计算结果如下。

例1 $f(x) = (x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2$，$x_0 = (2,2,2,2,2)^T$，$x^* = (1,1,1,1,1)^T$，$f^* = 0$，见表1。

表1: 例1计算结果的比较

IT	NA	FR	PRP	HS	SM
0	6.000(0)	6.000(0)	6.000(0)	6.000(0)	6.000(0)
10	3.100(-3)	2.800(-3)	2.300(-3)	4.200(-3)	1.730(-2)
20	1.965(-8)	3.092(-7)	7.365(-7)	1.252(-6)	1.082(-5)
25	3.291(-11)	9.576(-9)	9.855(-9)	2.171(-8)	1.056(-8)

例2　$f(x)=(x_1+10x_2)^4+5(x_3-x_4)^4+(x_2-2x_3)^4+10(x_1-x_4)^4$, $x_0=(2,2,-2,-2)^T$, $x^*=(0,0,0,0)^T$, $f^*=0$，见表2。

表2:　例2计算结果的比较

IT	NA	FR	PRP	HS	SM
0	2.381(5)	2.381(5)	2.381(5)	2.381(5)	2.381(5)
20	1.595(1)	1.362(2)	2.275(1)	2.791(1)	1.298(1)
30	2.860(-2)	1.542(0)	1.719(-1)	3.130(-2)	1.444(-1)
40	1.146(-4)	7.600(-3)	2.300(-3)	8.126(-5)	1.300(-3)

例3　$f(x)=(x_1-1)^2+(x_1-x_2)^2+(x_3-1)^2+(x_4-1)^4+(x_5-1)^6$, $x_0=(2,2,2,2,2)^T$, $x^*=(1,1,1,1,1)^T$, $f^*=0$，见表3。

表3:　例3计算结果的比较

IT	NA	FR	PRP	HS	SM
0	4.000(0)	4.000(0)	4.000(0)	4.000(0)	4.000(0)
10	3.360(-2)	1.200(-1)	1.650(-2)	1.380(-2)	1.090(-2)
20	3.632(-4)	6.662(-4)	8.066(-4)	3.845(-4)	6.656(-4)
30	9.628(-5)	1.434(-4)	2.185(-4)	2.523(-4)	2.810(-4)

例4　$f(x)=(1-x_1)^2+(1-x_{10})^2+\sum_{i=1}^{9}(x_i^2-x_{i+1})^2$, $x_0=(3,\cdots,3)^T$, $x^*=(1,\cdots,1)^T$, $f^*=0$，见表4。

表4:　例4计算结果的比较

IT	NA	FR	PRP	HS	SM
0	3.320(2)	3.320(2)	3.320(2)	3.320(2)	3.320(2)
20	6.189(-4)	1.020(-2)	4.070(-2)	8.300(-3)	1.050(-2)
30	9.862(-7)	4.226(-4)	2.786(-4)	5.727(-6)	7.075(-5)
50	1.738(-11)	7.345(-8)	2.537(-8)	5.205(-9)	1.042(-8)

例5　扩展Beale函数

$$f(x)=\sum_{i=1}^{n/2}\left\{\left[1.5-x_{2i-1}(1-x_{2i})\right]^2+\left[2.25-x_{2i-1}(1-x_{2i}^2)\right]^2+\left[2.625-x_{2i-1}(1-x_{2i}^3)\right]^2\right\},$$

$$n=2,4,\cdots,\quad x_0=(2,\cdots,2)^T,\quad x^*=(3,0.5,\cdots,3,0.5)^T,\quad f^*=0,$$

见表5。

表5: 例5计算结果的比较 ($n = 80$)

IT	NA	FR	PRP	HS	SM
0	1.467(4)	1.467(4)	1.467(4)	1.467(4)	1.467(4)
15	2.152(1)	2.727(1)	1.434(1)	2.293(1)	2.290(1)
20	3.759(0)	1.509(1)	2.206(0)	6.286(0)	3.659(0)
30	3.394(-1)	9.507(-1)	1.834(-1)	1.173(0)	3.075(-1)

参考文献：

[1] 袁亚湘，孙文瑜. 最优化理论与方法[M]. 北京：科学出版社，1997
Yuan Y X, Sun W Y. Optimization Theory and Methods[M]. Beijing: Science Press, 1997

[2] Miele A, Cantrell J W. Study on a memory gradient method for the minimization of functions[J]. JOTA, 1969, 3(6): 459-470

[3] 李敏，苏醒，时贞军. 求解非线性方程组的记忆梯度算法[J]. 工程数学学报，2009, 26(3): 563-566
Li M, Su X, Shi Z J. Memory gradient method for nonlinear equations[J]. Chinese Journal of Engineering Mathematics, 2009, 26(3): 563-566

[4] 孙清滢，谷亚丽，王长钰. 一个新的带误差项的记忆梯度算法[J]. 工程数学学报，2007, 24(5): 814-818
Sun Q Y, Gu Y L, Wang C Y. A new memory gradient method with errors[J]. Chinese Journal of Engineering Mathematics, 2007, 24(5): 814-818

[5] 时贞军. 非精确搜索下的超记忆梯度法[J]. 工程数学学报，2004, 21(3): 467-470
Shi Z J. On supermemory gradient method with inexact line searches[J]. Chinese Journal of Engineering Mathematics, 2004, 21(3): 467-470

[6] Shi Z J. A new super-memory gradient method for unconstrained optimization[J]. Advances in Mathematics, 2006, 35(3): 265-273

[7] 汤京永，时贞军. 一类全局收敛的记忆梯度法及其线性收敛性[J]. 数学进展，2007, 36(1): 67-75
Tang J Y, Shi Z J. A class of global convergent memory gradient methods and its linear convergence rate[J]. Advances in Mathematics, 2007, 36(1): 67-75

[8] 汤京永，时贞军. 一类新的强Wolfe线性搜索下的记忆梯度法[J]. 曲阜师范大学学报，2005, 31(2): 24-29
Tang J Y, Shi Z J. A new class of memory gradient methods under strong Wolfe line search[J]. Journal of Qufu Normal University, 2005, 31(2): 24-29

A New Memory Gradient Method and its Convergence

TANG Jing-yong[1,2], DONG Li[1]

(1- College of Mathematics and Information Science, Xinyang Normal University, Xinyang 464000;
2- Department of Mathematics, Shanghai Jiaotong University, Shanghai 200240)

Abstract: The memory gradient methods for unconstrained optimization problems were investigated. A new algorithm is presented which uses the current and previous one-step iterative information to generate a decent direction, and uses an Armijo linear search to determine the step-size. The new method converges globally and it has a linear convergence rate under some mild conditions. Moreover, the method avoids the computation and storage of some matrices. It is suitable for solving large scale optimization problems. Experimental results show that the new method is efficient in practical computation.

Keywords: unconstrained optimization; memory gradient method; global convergence; linear convergence rate

Received: 15 Aug 2008. **Accepted:** 17 Sep 2009.

Foundation item: The National Natural Science Foundation of China (10571109); the Youth Foundation of Xinyang Normal University (200946, 200947).